Don't Miss Out!

Eekhout, I.

2015

Chapter 4

Multiple imputation at the item level when the number of items is very large

Under review: Eekhout, I., de Vet, H.C.W., de Boer, M.R., Twisk, J.W.R., Heymans, M.W. Passive imputation of missing values in studies with many multi-item questionnaire outcomes. Quality of Life Research.

Abstract

Previous studies showed that missing data in multi-item questionnaires should be handled by multiple imputation. However, when many questionnaires are used the number of item variables with missing values will become too large to reliably estimate imputations. Passive imputation methods have been developed to combine variables in the imputation model to reduce information, which has never been studied before in the situation of missing item scores. In a simulation study we compared five methods as part of the multiple imputation procedure in RCTs with complete-case analysis when item scores were made missing in five different multi-item questionnaires. Method 1 and 2 used passive imputation, which updated the questionnaire total scores from imputed item variables between imputations, method 3 used parcel summary scores of the items, method 4 used all items at once and method 5 directly imputed the total scores. Descriptive statistics of questionnaire total scores and treatment coefficient estimates from linear regression were compared to 'true' parameters on bias, mean squared error and coverage. Passive imputation and using parcel summaries showed a standardized bias of less than 10%, while imputing the total score directly a standardized bias of over 60% for the questionnaire total scores. The sample size for imputing total scores needs to be at least 23% larger to attain the same mean squared error in regression coefficients compared to passive imputation. Item imputations are most valid when passive imputation or parcel summary scores are used. These methods are therefore recommended for missing data in multi-item questionnaires.

Keywords: multi-item questionnaires, missing data, multiple imputation, passive imputation, large survey studies

Introduction

Many epidemiological and medical studies use multiple questionnaires to measure patient characteristics or disease outcome. These questionnaires consist of several items which are usually measured at several time-points resulting in a dataset with many variables representing the items. Subsequently, the item scores for each questionnaire are summed up and the total scores can be used in the analyses as predictors, covariates or outcomes. However, often these questionnaires contain missing data on the item scores, which impairs the calculation of the total scores.

An advanced method to handle missing data is multiple imputation (MI) (Rubin, 1987; Schafer, 1997). In MI the missing values are replaced by multiple plausible values resulting in multiple copies of the dataset, each with different imputed values for the missing entries. The plausible values are estimated in an imputation model utilizing regression models to predict and replace missing values based on the observed data. The data analysis is then performed on each imputed dataset, resulting in multiple sets of results. In the end, these sets of results are pooled into one final result. Multiple imputation can either be applied to the questionnaire item scores before the total score is calculated, or to the total scores directly, in which case the total scores are incomplete when one or more items are missing. From previous studies we know that it is most advantageous to handle missing data in multi-item questionnaires at the item level (Eekhout et al., 2014; Gottschall, West, & Enders, 2012).

Since multiple imputation involves using regression models to estimate the imputed values, the rules and assumptions for regression analyses also apply to multiple imputation. One limitation of regression analysis is that the number of independent variables cannot be too large, which can be a problem when all items are included at once in the multiple imputation model (Green, 1991; Hardt, Herke, & Leonhart, 2012). By running the imputation process for each questionnaire or outcome separately, information is lost because the questionnaires might be related to each other. It is recommended to include all possible information in the imputation model and therefore, it is most informative to incorporate all questionnaires at once to deal with missing data (Collins, Schafer, & Kam, 2001).

A possible solution to avoid the problem of having imputation models that are too large is to use total scores of questionnaires in the imputation model as predictors for the missing item scores instead of using only item scores. This seems like a straightforward solution, however, these total scores often contain missing values, which are also caused by missing item scores. A solution is to adapt the imputation process in such a way that the total score will be calculated after each imputation run (i.e., iteration) using the imputed item scores. This is possible by an application called passive imputation. Passive imputation can be used to make sure that a derived variable (e.g., a questionnaire total score calculated by the sum of the item scores) always depends on the most recently generated item imputations in the original data (van Buuren, 2012). Accordingly, between imputation iterations, the total score is updated from the most recently imputed item scores, which is the passive part of the imputation. Furthermore, during the iteration process the total scores of the questionnaire cannot be used as a predictor for items of that specific questionnaire, but only as a predictor for the items of other questionnaires. Passive imputation seems perfectly designed to handle missing values in different items using several different questionnaires, with the benefit of maintaining the imputation model without the problem of a large number of variables in the imputation model.

Passive imputation in the context of interaction variables (i.e., ratios of variables) has been studied previously and was found to result in biased regression estimates (Morris, White, Royston, Seaman, & Wood, 2014; Von Hippel, 2009). The application of passive imputation for questionnaires was briefly proposed by Van Buuren (2010), however the validity of this method under different data situations has not been studied before. This study therefore evaluated two procedures of the passive imputation method for the imputation of item scores in simulated data. Furthermore, these passive imputation methods were compared to a practical method that imputes the items by using a parcel summary score of the other questionnaires as predictors, which can be applied in any software package. Finally, these methods were also compared to imputing the item scores, imputing total scores directly and to a complete-case analysis (CCA). The latter two methods are mostly used in practice (Eekhout, de Boer, Twisk, de Vet, & Heymans, 2012).

Methods

Simulation study design

We simulated data for five questionnaires (Q1 to Q5). The first simulated questionnaire (Q1) contained 5 items, the second (Q2) and third (Q3) questionnaires contained 10 items and the fourth (Q4) and fifth (Q5) 15 items. All items were measured on a five-point Likert scale. We simulated a randomized controlled trial situation with a pre and a post measurement for the questionnaires, because this is a frequently applied study design in epidemiological studies. Additionally, two baseline continuous covariates were simulated and a random dichotomous treatment variable. This resulted in a total of 55 items measured at two time-points, two time-invariant continuous covariates and one dichotomous time-invariant covariate (i.e., treatment). For the simulation we used a predefined treatment effect of 0.50 for the total scores. Furthermore, we varied between two sample size conditions: 150 subjects and 250 subjects per simulated dataset separated in two equal treatment groups. We generated 1000 samples in each sample size condition. The complete data samples were used as a reference to compare the performance of the missing data methods to. The complete samples were created in Mplus (Muthén & Muthén, 1998-2012). Subsequently, missing data in the items were generated in all questionnaires by the missing at random mechanism (Rubin, 1976). For 10-25% of the subjects only some items were incidentally made missing within subjects (i.e., < 75% of the items) and for 0-12% of the subjects a whole questionnaire was made missing (i.e., > 75% of the items were missing). These percentages varied per questionnaire. The probability of missing an entire questionnaire was larger after treatment than at baseline (i.e., 0-6% at baseline and 6-12% post-treatment). That way a realistic data situation was simulated where some people skipped some questionnaire items and other people didn't fill out an entire questionnaire. The overall percentage of subjects with missing data was simulated to be 30%. The missing data in the samples were generated in R statistical software (R Core Development Team, 2014).

Compared multiple imputation methods

In the simulated datasets, the missing data were handled with multiple imputation by multivariate imputation by chained equations (MICE) (van Buuren & Groothuis-Oudshoorn, 2011). The ordinal Likert items of the questionnaires in our simulated data were imputed with the predictive mean matching method, which assumes normality but was shown to work for ordinal items as well (Eekhout, et al., 2014). There are several options for specification of the imputation model when questionnaire items and total scores at multiple time-points (e.g., baseline and post-treatment) are involved. We compared four different imputation models that were targeted at imputing missing item scores and one model that imputed the total scores of the questionnaires directly and a CCA. Each imputation model included the treatment group variable and the two other covariates. The following methods were compared:

Method 1: Passive imputation A (M1-Passive)

For this passive imputation procedure the imputation model consisted of the following variables: the item variables of a questionnaire assessed at a certain time-point, the item variables from this questionnaire assessed at other time-points and the total scores of the other questionnaires at both time-points. With this method the total scores are updated after each imputation iteration by the imputed items from the previous iteration. Then the updated total scores become the predictors for the item variables that contain missing values during the next iterations. For the smaller questionnaire (Q1) the imputation model contained 21 variables (i.e., 5 items at baseline, 5 items post-treatment for Q1, 8 total scores for Q2-Q5 at baseline and post-treatment, the treatment group variable and the two covariates), for Q2 & Q3

31 variables and for the larger questionnaires (Q4 & Q5) 41 variables.

Method 2: Passive imputation B (M2-Passive)

The second imputation model also included passive imputation, where the items for a questionnaire at a certain time-point, the total score of that questionnaire at the other time-point and the total scores of all other questionnaires at both time-points were included in the imputation model. Accordingly, the method includes fewer variables in the imputation model at once compared with M1-Passive. The model contained 17 variables for the smaller questionnaire (Q1) (i.e., the 5 items of Q1 at baseline, the total score for Q1 post-treatment, 8 total scores for Q2-Q5 at baseline and post-treatment, the treatment group variable and the two covariates), 22 variables for Q2 & Q3 and 27 variables for the larger questionnaires (Q4 & Q5).

Method 3: Parcel summary model (M3-Parcel)

The third imputation model included the average of the available items of the other questionnaires as predictors for the missing item scores of a questionnaire. The average over the available items is a parcel summary score of the item information, which was calculated once for each questionnaire prior to the start of the imputation process. This method does not use passive imputation, but the same parcel summary scores were used as predictors in the imputation model for each questionnaire. In M3-Parcel the imputation for each separate questionnaire was done independently and we merged the resulting imputed datasets after the imputation process was completed. In this method 21 variables were in the imputation model for the smaller questionnaire (Q1) (i.e., 5 items at baseline, 5 items post-treatment for Q1, 8 parcel summary scores for Q2-Q5 at baseline and post-treatment, the treatment group variable and the two covariates), for Q2 & Q3 31 variables and 41 variables for the larger questionnaires (Q4 & Q5).

Method 4: All item scores (M4-Items)

In the fourth imputation model all items in the dataset were included at once. Consequently, 110 items, the treatment group variable and the two covariates were all entered in the imputation model at once. This model was expected to encounter convergence problems, especially in the smaller sample size condition, but it was included as a comparison.

Method 5: Total scores (M5-TS)

The fifth imputation model was targeted at total scores directly. The total scores were computed prior to imputation and were incomplete when one or more items were missing. The imputation model included only the total scores of all questionnaires Q1 to Q5 at both time-points, the treatment group variable and the two covariates (i.e., 13 variables). The item scores were ignored in this procedure.

Method 6: Complete-case analysis (M6-CCA).

We also performed a CCA on the data where the total scores were also left incomplete when one or more item scores were missing. In the CCA only the subjects with completely observed data were included in the data analysis. This resulted in a reduction of the sample size of 30%.

As described above, our simulated data contained subjects who had incidental missing item scores as well as subjects that missed data on an entire questionnaire. To accommodate this, we applied the method 1 to 4 as follows. For the subjects that had the entire questionnaire missing (i.e., >75% of the item scores missing within a subject), the total scores were imputed directly. Subsequently, for the subjects that had less than 75% of the item scores missing (i.e., incidental missing item scores within subjects), the item scores were imputed according to the method 1 to 4. Then after the imputation, but before analysis, we selected the total score from the imputed items for the people who missed only some items of a questionnaire and for the people that missed the entire questionnaire, we selected the directly imputed total score.

Analyses

The imputed and thus complete data were analyzed by linear regression analysis using the questionnaire total score after treatment as the outcome and the treatment variable as covariate, adjusted for the baseline measurement of the questionnaire. We analyzed each of the five questionnaires (the outcome variables) separately and we were interested in the treatment coefficients. Furthermore, we computed the means of the questionnaire total scores at baseline and post-treatment. The results for all methods were compared to the complete data results using bias, mean squared error (MSE) and coverage. The bias was evaluated by examining the standardized bias, which reflects the bias relative to the overall uncertainty in the sampling (Collins, et al., 2001). The standardized bias was calculated by

$$Standardized\ bias = \frac{\bar{\bar{\beta}} - \bar{\beta}_c}{sd(\beta_c)}\ 100\%$$

where $\bar{\bar{\beta}}$ is the average parameter estimate (e.g., the questionnaire mean or the treatment regression coefficient) obtained from the estimates in the simulated datasets after the missing data method was applied, $\bar{\beta}_c$ is the average true parameter of the simulated complete reference data and $sd(\beta_c)$ is the standard deviation of the estimates of the complete data. The MSE represents the precision and accuracy of the estimates and was calculated by

$$MSE = \left(\bar{\hat{\beta}} - \bar{\beta}_c\right)^2$$

The coverage was calculated by the percentage of times the average complete data estimate $\bar{\beta}_c$ was within the 95% confidence interval of the estimated parameters $\hat{\beta}_i$. For a 95% confidence interval, the coverage rate should be at 95%. Coverage rates higher than 95% indicate that the method might be too conservative, and a lower coverage rate suggests higher than expected type I error (Burton & Altman, 2004). All imputations and analyses were performed in R statistical software (R Core Development Team, 2014). A detailed manual on how to apply the imputation methods is available by the first author.

Results

In Figure 4.1 the standardized bias of the regression coefficients for the treatment coefficient for each questionnaire outcome are presented for the missing data methods at both sample size conditions.

For the sample size of 150, the method that included all items in the imputation model (M4-Items), could not be computed because of an excess of items in the imputation model. Overall we can observe that all methods including the CCA resulted in small standardized bias for the regression coefficients. In both sample size conditions the standardized bias in the regression coefficients (i.e., in treatment effect) was smaller than 10%. In the small sample size condition, the methods applied to the item scores (M1-Passive, M2-Passive & M3-Parcel) performed slightly better than the method applied to total scores (M5-TS); however, differences are very small (Figure 4.1a). In the larger sample size condition, multiple imputation applied to the total scores (M5-TS) had a standardized bias comparable to the methods applied to the item scores (Figure 4.1b).

The MSE of the regression coefficient estimates increased when the number of items per questionnaire increased (Figure 4.2). Furthermore, we can observe that the methods that imputed the item scores (M1 to M4) had a smaller MSE (i.e., more precision and accuracy) than the method that was applied to the total scores (M5-TS) and this difference increased when the number of items per questionnaire was increasing. For the smaller questionnaire (Q1) at n=150 the MSE was 1.22 for both passive imputation methods (M1-Passive & M2-Passive) and for imputing total scores (M5-TS) 1.35. The ratio of these MSEs (1.35/1.22=1.11) indicates the sample size increase required for imputing total scores to attain the same precision as the passive imputation methods, which is 11%. For the larger questionnaires (e.g., Q5) at n=150 the MSE of passive imputation was 9.43 and for imputing total scores 12.36,

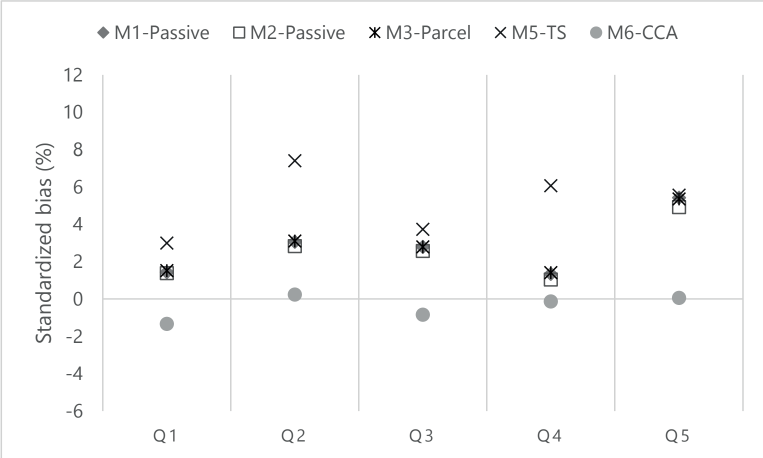

Figure 4.1a. Standardized bias of the regression coefficients for treatment effect for each of the five questionnaires for the missing data methods (n=150). M4-Items could not be performed for n=150.

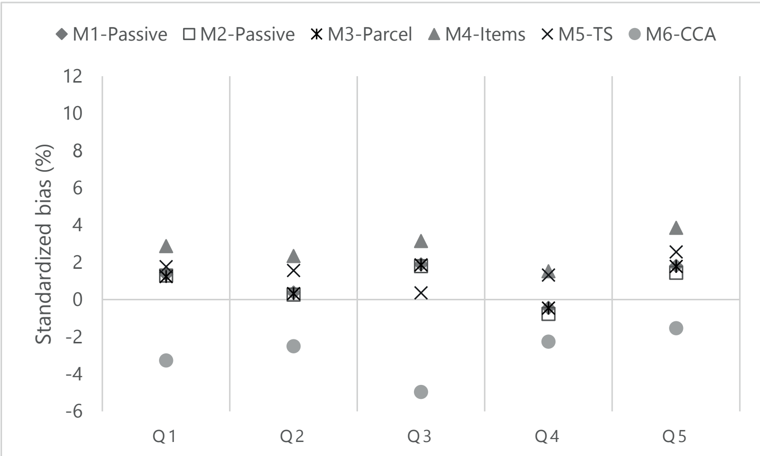

Figure 4.1b. Standardized bias of the regression coefficients for treatment effect for each of the five questionnaires for the missing data methods (n=250).

accordingly the ratio was (12.36/9.43) 1.31, which indicates a required sample size increase of 31% for imputing total scores to reach the same precision as passive imputation. On average over all questionnaires (Q1 to Q5) the ratio of the MSE of passive imputation to the MSE of imputing total scores was 1.23, which means that the sample size should increase by at least 23% for imputing total scores to attain the same precision as passive imputation. The MSE of CCA was even worse (i.e., larger); the average ratio of the MSE of passive imputation to the MSE of CCA was 1.33, which indicates required a sample size increase by 33% for CCA to achieve the same precision as passive imputation. The coverage rates of the regression coefficient estimates were satisfactory for all methods for both sample size conditions (data not shown).

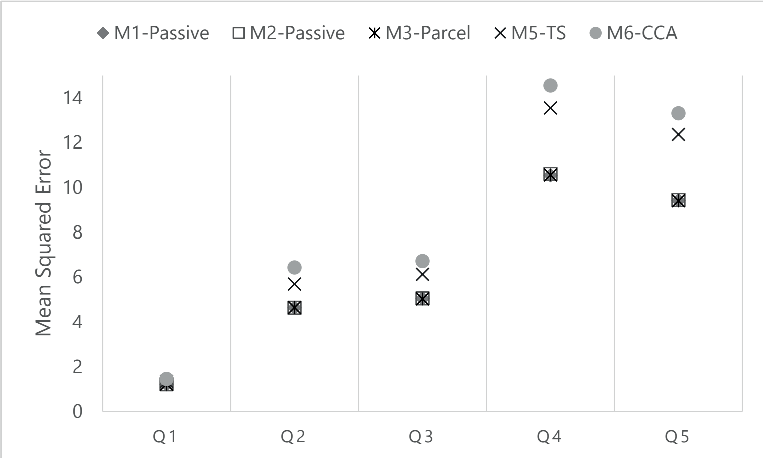

Figure 4.2a. Mean Squared Error (MSE) for the regression coefficients for treatment effect of the analysis of the five questionnaires for the missing data methods (n=150). M4-Items could not be performed for n=150.

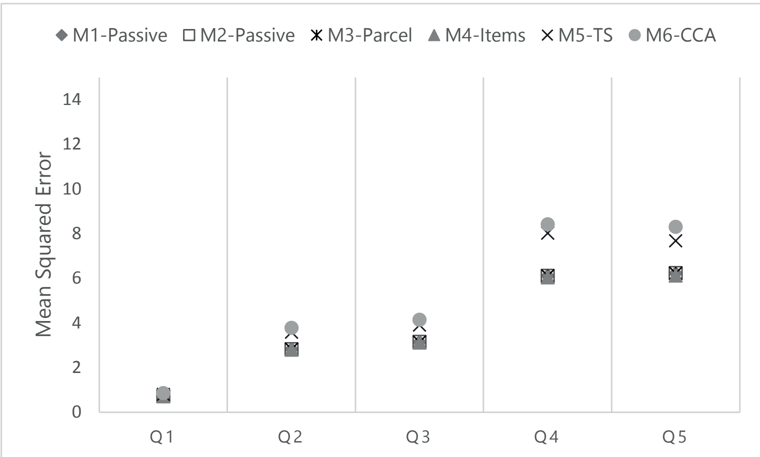

Figure 4.2b. Mean Squared Error (MSE) for the regression coefficients for treatment effect of the analysis on the five questionnaires for the missing data methods (n=250)

The coverage rates of the regression coefficient estimates were satisfactory for all methods for both sample size conditions (data not shown).

In Figure 4.3 the standardized bias of the means of the questionnaire total scores are presented for each method. In both sample size conditions we can observe that the passive imputation methods (M1-Passive & M2-Passive) performed best and were least biased. The parcel summary method (M3-Parcel) performed similarly to the passive imputation methods. In the larger sample size conditions (n=250), the method that included all the items in the imputation model at once (M4-Items), also performed similarly. Imputing the total score directly (M5-TS) performed less favorable, i.e., larger standardized bias. CCA performed worst and showed bias larger than 1 standard error (i.e., 100%) in both sample size conditions.

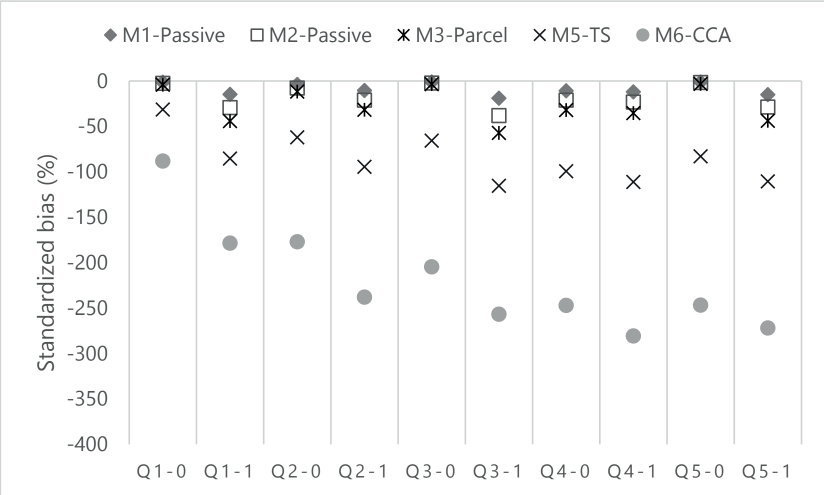

Figure 4.3a. Standardized bias of questionnaire total score mean at baseline (0) and after treatment (1) for each questionnaire for the compared missing data, n=150. M4-Items could not be performed for n=150.

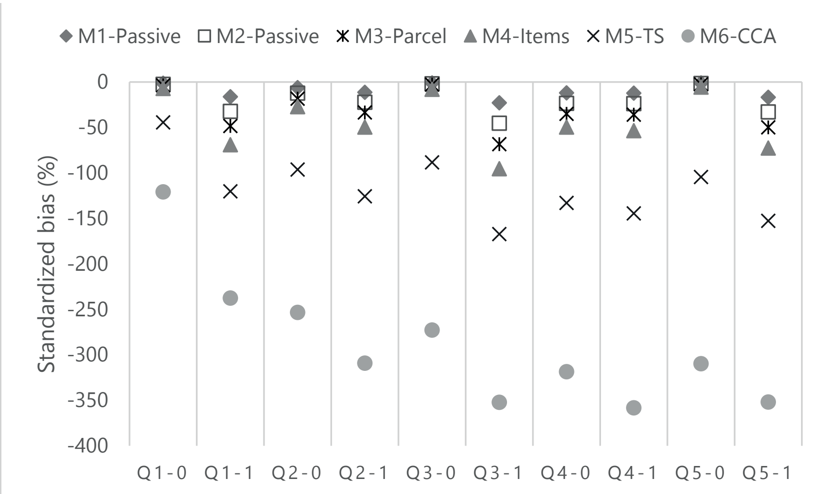

Figure 4.3b. Standardized bias of questionnaire total score mean at baseline (0) and after treatment (1) for each questionnaire for the compared missing data, n=250.

In Figure 4.4 the average MSE of the questionnaire total scores for each imputation method are presented. We can observe the trend that the methods that impute the item scores (M1-M4) had the lowest MSE and performed almost similarly. The imputation of the total scores (M5-TS) resulted in about one-and-a-half times larger MSE and the CCA is even less precise.

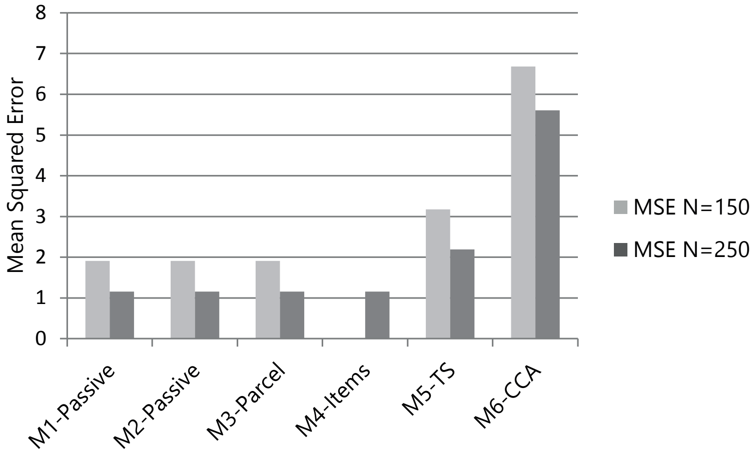

Figure 4.4. MSE of questionnaire total score means averaged for all questionnaires at baseline and after treatment for the missing data methods, at the two sample size conditions (n=150 & n=250). M4-Items could not be performed for n=150.

Discussion

In this paper we compared possible solutions for the problem of missing data when data from many questionnaires are used in one study. To our knowledge, the proposed new methods that use passive imputation (M1-Passive & M2-Passive) had not been validated yet. The results of our simulation show that using passive imputation to impute the item scores of multiple questionnaires is a valid solution for missing item score data in questionnaires. In fact, all methods in which the item scores are imputed by using the total scores as predictors in the imputation model, including the method where a parcel summary score of the items from the other questionnaires was used, performed better than handling the missing data directly at the total score level. This illustrates the importance of including the item scores in the imputation model. Even when the most optimal solution (i.e., including all available item information) is not viable, it is important to apply an advanced missing data method that incorporates the item scores. In this study we found differences in performance for the methods when they were applied to estimate the group means or the treatment effects. In the analysis of treatment effects, CCA was unbiased, though the precision reflected in the MSE was much better for the imputation methods. For the estimation of the group means the imputation methods performed superiorly compared to a CCA on all evaluation measures (i.e., bias and MSE).

From previous simulation studies we know that it is most beneficial to handle missing data in multi-item questionnaires at the item level (Eekhout, et al., 2014). However, the most appropriate method, i.e., the imputation model where solely the item score information was incorporated (M4-Items), is only viable with a sufficiently large sample size. For that reason, in many situations a solution that includes a smaller imputation model is necessary. The methods that incorporate the item score information most optimally were the passive imputation methods that impute the item scores by using the total scores from the other questionnaires as predictors in the imputation model (M1-Passive & M2-Passive). In these methods, the total scores were updated by the imputed items within the imputation process. That way the most optimal available information was used at every stage of the imputation process. The performance of the passive imputation methods was best on all accounts (i.e., descriptive statistics and regression estimates of the questionnaire data). For that reason this strategy is advised. We compared two different ways to apply these methods. The first strategy (M1-Passive) was to include the item scores for both time-points of one questionnaire together in one batch of imputations. The second strategy was to impute the item scores for the baseline of one questionnaire separately from the post-treatment item scores of that questionnaire, so for each time-point separately. Both strategies performed equally well, so the only preference is related to the number of variables that is included in the imputation model, which might be preferred to be kept lower (i.e., M2-Passive).

Previous studies evaluated the performance of passive imputation when the imputation model contained ratios of variables (Morris, et al., 2014), interactions between variables or squared variables (Von Hippel, 2009). For these kinds of composed variables, the variables are separately imputed in their raw form, and then their composed values are calculated between each iteration of imputations as an update. In that case, the raw variables are imputed in the imputation model, while the analysis model contains the transformed variables (i.e., ratios, interactions or squares). These studies concluded that the use of passive imputation for these kinds of variables can result in biased parameter estimates, because the covariance between the predictor variables and the outcome variable in the imputation model is different from the covariance between the predictors and the outcome in the analysis model. For that reason, these studies advised to transform the variables prior to the imputation. In the present study we did not use any transformations and therefore the covariance matrix in the imputation model and the analysis model are compatible.

Passive imputation is available in the MI procedure in STATA, the MICE package in R and S-PLUS and in IVEware in SAS. The application of this method requires an advanced level of programming and might therefore not be feasible in daily practice for many researchers. For that reason we included an imputation strategy that can be applied in most basic statistical programs that include a multiple imputation option (e.g., SPSS). This method (M3-Parcel), which uses the average of the available items from the other questionnaires to impute the items of one questionnaire, performed satisfactory with regard to most parameters as well. Only the average total scores were biased for this method, but the coefficient estimates for treatment effect and the precision of these estimates were adequate and better than imputing the total scores directly.

Strengths and weaknesses

The strength of this simulation study is that we simulated realistic situations with both items and total scores missing. Furthermore, the design of including several questionnaires in one study is very common in epidemiology. Also different lengths of questionnaires were included with two different sample size conditions. That way the sensitivity of the missing data methods for several data aspects was checked. The tested methods have not been validated before in previous studies and the results of this study underpin the good performance of the newly proposed methods.

A possible weakness of this study might be that missing data methods that are advised in most user-manuals of multi-item questionnaires were not included in the comparison. These advised methods are mostly single imputation methods, for example replacing the missing value with the mean score (Lambert, Lunnen, Umphress, Hansen, & Burlingame, 1994; Ware, Kosinski, & Keller, 1994). However, from a previous simulation study we know that these methods do not perform well and are not recommended to be applied in any missing data situation (Eekhout, et al., 2014). Furthermore, a limited amount of data conditions was compared in this study. For example, we did not vary the total number of items in the entire dataset. However, by varying the sample size we aimed to simulate a situation where the ratio of number of items and sample size varied, which is the most important reason for one of these multiple imputation methods to fail. The amount of missing data was not varied either. However we aimed to simulate a realistic amount of missing data and we expect our methods to perform similarly with less missing data. Moreover, in our previous simulation we found that multiple imputation of item scores remains to perform well up to conditions with 75% missing item scores within 75% of the subjects (Eekhout, et al., 2014).

All compared multiple imputation methods (M1-M5) are advanced methods to handle missing data. For that reason it was expected that the performance of all methods would be satisfactory to some extent, for example on coverage values. However, this simulation showed that the new methods outperform the current solution of imputing questionnaire data when many questionnaire items are involved (i.e., M5-TS), especially on precision but also on accuracy. In general we advise to include as much item score information as possible in handling the missing data at the item score level of a multi-item questionnaire. It is best to impute the item scores by using the passive imputation procedure.

References

Burton, A., & Altman, D. G. (2004). Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. British Journal of Cancer, 91(1), 4-8.

van Buuren, S. (2010). Item Imputation Without Specifying Scale Structure. Methodology, 6(1), 31-36.

van Buuren, S. (2012). Flexible Imputation of Missing data. New York: Chapman & Hall/CRC.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychological Methods, 6(4), 330-351.

Eekhout, I., de Boer, R. M., Twisk, J. W., de Vet, H. C., & Heymans, M. W. (2012). Missing data: a systematic review of how they are reported and handled. Epidemiology, 23(5), 729-732.

Eekhout, I., De Vet, H. C. W., Twisk, J. W. R., Brand, J. P. L., De Boer, M. R., & Heymans, M. W. (2014). Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. Journal of Clinical Epidemiology, 67(3), 335-342.

Gottschall, A. C., West, S. G., & Enders, C. K. (2012). A Comparison of Item-Level and Scale-Level Multiple Imputation for Questionnaire Batteries. Multivariate Behavioral Research, 47(1), 1-25.

Green, S. B. (1991). How Many Subjects Does It Take To Do A Regression Analysis. Multivariate Behavioral Research, 26(3), 499-510.

Hardt, J., Herke, M., & Leonhart, R. (2012). Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. BMC Med Res Methodol, 12, 184.

Lambert, M. J., Lunnen, K., Umphress, V., Hansen, N., & Burlingame, G. M. (1994). Administration and scoring manual for the Outcome Questionnaire (OQ-45.1). Salt Lake City: IHC Center for Behavioral Healthcare Efficacy.

Morris, T. P., White, I. R., Royston, P., Seaman, S. R., & Wood, A. M. (2014). Multiple imputation for an incomplete covariate that is a ratio. Statistics in Medicine, 33(1), 88-104.

Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3), 581-592.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: J. Wiley & Sons.

Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data. London, UK: Chapman & Hall.

R Core Development Team (2014). R: A language and environment for statistical computing. Vienna, Austria.

Von Hippel, P. T. (2009). How to impute interactions, squares, and other transformed variables. Sociological Methodology, 39(1), 265-291.

Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1994). SF-36 Physical and Mental Health Summary Scales: A User’s Manual. Boston, MA: Health Assessment Lab.